



Anthropic

Anthropic PBC is an American artificial intelligence (AI) startup company founded in 2021. Anthropic has developed a family of large language models (LLMs) named Claude as a competitor to OpenAI's ChatGPT and Google's Gemini.^[5] According to the company, it researches and develops AI to "study their safety properties at the technological frontier" and use this research to deploy safe models for the public.^{[6][7]}

Anthropic was founded by former members of OpenAI, including siblings Daniela Amodei and Dario Amodei.^[8] In September 2023, Amazon announced an investment of up to \$4 billion, followed by a \$2 billion commitment from Google in the following month.^{[9][10][11]}

History

Founding and early development (2021–2022)

Anthropic was founded in 2021 by seven former employees of OpenAI, including siblings Daniela Amodei and Dario Amodei, the latter of whom served as OpenAI's Vice President of Research.^{[12][13]}

In April 2022, Anthropic announced it had received \$580 million in funding,^[14] including a \$500 million investment from FTX under the leadership of Sam Bankman-Fried.^{[15][3]}

In the summer of 2022, Anthropic finished training the first version of Claude but did not release it, mentioning the need for further internal safety testing and the desire to avoid initiating a potentially hazardous race to develop increasingly powerful AI systems.^[16]

Legal and strategic partnerships (2023)

In February 2023, Anthropic was sued by Texas-based Anthrop LLC for the use of its registered trademark "Anthropic A.I."^[17] On September 25, 2023, Amazon announced a partnership with Anthropic, with Amazon becoming a minority stakeholder by initially investing \$1.25 billion, and planning a total

Anthropic PBC

ANTHROPIC

Logo

Company type	<u>Private</u>
Industry	<u>Artificial intelligence</u>
Founded	2021
Founders	<u>Dario Amodei</u> <u>Daniela Amodei</u> <u>Jared Kaplan</u> ^[1] <u>Jack Clark</u> ^[2] <u>Ben Mann</u> ^[3]
Headquarters	<u>San Francisco, California</u> , <u>U.S.</u>
Products	<u>Claude</u>
Number of employees	c. 800 (2024) ^[4]
Website	<u>anthropic.com</u> (<u>https://www.anthropic.com/</u>)



Dario Amodei, co-founder

investment of \$4 billion.^[9] As part of the deal, Anthropic would use Amazon Web Services (AWS) as its primary cloud provider and make its AI models available to AWS customers.^{[9][18]} The next month, Google invested \$500 million in Anthropic, and committed to an additional \$1.5 billion over time.^[11]

Major investments and acquisitions (2024)

In March 2024, Amazon maxed out its potential investment from the agreement made in the prior year by investing another US\$2.75 billion into Anthropic, completing its \$4 billion investment.^[10]

In November 2024, Amazon announced a new investment of \$4 billion in Anthropic (bringing its total investment to \$8 billion), including an agreement to increase the use of Amazon's AI chips for training and running Anthropic's large language models.^[19]

In 2024, Anthropic attracted several notable employees from OpenAI, including Jan Leike, John Schulman, and Durk Kingma.^[20]

Business structure

According to Anthropic, the company's goal is to research the safety and reliability of artificial intelligence systems.^[7] The Amodei siblings were among those who left OpenAI due to directional differences.^[13]

Anthropic incorporated itself as a Delaware public-benefit corporation (PBC), which enables directors to balance the financial interests of stockholders with its public benefit purpose.^[21]

Anthropic's "Long-Term Benefit Trust" is a purpose trust for "the responsible development and maintenance of advanced AI for the long-term benefit of humanity". It holds Class T shares in the PBC which allow it to elect directors onto Anthropic's board.^{[22][23]} As of April 2025, the members of the Trust are Neil Buddy Shah, Kanika Bahl and Zach Robinson.^[24]



Anthropic's corporate structure

Investors include Amazon.com for \$8B,^[19] Google for \$2B,^[11] and Menlo Ventures for \$750M.^[25]

Key employees

- Dario Amodei: Co-founder and chief executive officer^[8]
- Daniela Amodei: Co-founder and President^[8]
- Mike Krieger: Chief Product Officer^{[26][27]}
- Jan Leike: ex-OpenAI alignment researcher^[28]

Projects

Claude

Claude incorporates "Constitutional AI" to set safety guidelines for the model's output.^[29] The name, "Claude", was chosen either as a reference to mathematician Claude Shannon, or as a male name to contrast the female names of other A.I. assistants such as Alexa, Siri, and Cortana.^[3]

Anthropic initially released two versions of its model, Claude and Claude Instant, in March 2023, with the latter being a more lightweight model.^{[30][31][32]} The next iteration, Claude 2, was launched in July 2023.^[33] Unlike Claude, which was only available to select users, Claude 2 is available for public use.^[34]

Claude 3 was released in March 2024, with three language models: Opus, Sonnet, and Haiku.^{[35][36]} The Opus model is the largest. According to Anthropic, it outperformed OpenAI's GPT-4 and GPT-3.5, and Google's Gemini Ultra, in benchmark tests at the time. Sonnet and Haiku are Anthropic's medium- and small-sized models, respectively. All three models can accept image input.^[35] Amazon has added Claude 3 to its cloud AI service Bedrock.^[37]

In May 2024, Anthropic announced the Claude Team plan, its first enterprise offering for Claude, and Claude iOS app.^[38]

In June 2024, Anthropic released Claude 3.5 Sonnet, which demonstrated significantly improved performance on benchmarks compared to the larger Claude 3 Opus, notably in areas such as coding, multistep workflows, chart interpretation, and text extraction from images. Released alongside 3.5 Sonnet was the new Artifacts capability in which Claude was able to create code in a dedicated window in the interface and preview select code in real time such as websites or SVGs.^[39]

In October 2024, Anthropic released an improved version of Claude 3.5, along with a beta feature called "Computer use", which enables Claude to take screenshots, click, and type text.^[40]

In November 2024, Palantir announced a partnership with Anthropic and Amazon Web Services to provide U.S. intelligence and defense agencies access to Claude 3 and 3.5. According to Palantir, this was the first time that Claude would be used in "classified environments".^[41]

In December 2024, Claude 3.5 Haiku was made available to all users on web and mobile platforms.^[42]

In February 2025, Claude 3.7 Sonnet was introduced to all paid users. It is a "hybrid reasoning" model (one that responds directly to simple queries, while taking more time for complex problems).^{[43][44]}

Constitutional AI

According to Anthropic, Constitutional AI (CAI) is a framework developed to align AI systems with human values and ensure that they are helpful, harmless, and honest.^{[12][45]} Within this framework, humans provide a set of rules describing the desired behavior of the AI system, known as the "constitution".^[45] The AI system evaluates the generated output and then adjusts the AI models to better fit the constitution.^[45] The self-reinforcing process aims to avoid harm, respect preferences, and provide true information.^[45]

Some of the principles of Claude 2's constitution are derived from documents such as the [1948 Universal Declaration of Human Rights](#) and Apple's terms of service.^[33] For example, one rule from the UN Declaration applied in Claude 2's CAI states "Please choose the response that most supports and encourages freedom, equality and a sense of brotherhood."^[33]

Interpretability research

Anthropic also publishes research on the [interpretability](#) of [machine learning](#) systems, focusing on the [transformer](#) architecture.^{[12][46][47]}

Part of Anthropic's research aims to be able to automatically identify "features" in [generative pretrained transformers](#) like Claude. In a [neural network](#), a feature is a pattern of neural activations that corresponds to a concept. In 2024, using a compute-intensive technique called "[dictionary learning](#)", Anthropic was able to identify millions of features in Claude, including for example one associated with the [Golden Gate Bridge](#). Enhancing the ability to identify and edit features is expected to have significant safety implications.^{[48][49][50]}

In March 2025, research by Anthropic suggested that multilingual LLMs partially process information in a conceptual space before converting it to the appropriate language. It also found evidence that LLMs can sometimes plan ahead. For example, when writing poetry, Claude identifies potential rhyming words before generating a line that ends with one of these words.^{[51][52]}

U.S. military and intelligence

Anthropic partnered with [Palantir](#) and [Amazon Web Services](#) in November 2024 to provide the Claude model to U.S. intelligence and defense agencies.^[53] Anthropic's CEO [Dario Amodei](#) said about working with the U.S. military:

The position that we should never use AI in defense and intelligence settings doesn't make sense to me. The position that we should go gangbusters and use it to make anything we want — up to and including doomsday weapons — that's obviously just as crazy. We're trying to seek the middle ground, to do things responsibly.^[54]

Lawsuit

On October 18, 2023, Anthropic was sued by [Concord](#), [Universal](#), [ABKCO](#), and other music publishers for, per the complaint, "systematic and widespread infringement of their copyrighted song lyrics."^{[55][56][57]} They alleged that the company used copyrighted material without permission in the form of song lyrics.^[58] The plaintiffs asked for up to \$150,000 for each work infringed upon by Anthropic, citing infringement of copyright laws.^[58] In the lawsuit, the plaintiffs support their allegations of copyright violations by citing several examples of Anthropic's Claude model outputting copied lyrics from songs such as [Katy Perry's "Roar"](#) and [Gloria Gaynor's "I Will Survive"](#).^[58] Additionally, the plaintiffs alleged that even given some prompts that did not directly state a song name, the model responded with modified lyrics based on original work.^[58]

On January 16, 2024, Anthropic claimed that the music publishers were not unreasonably harmed and that the examples noted by plaintiffs were merely bugs.^[59]

In August 2024, a class-action lawsuit was filed against Anthropic in California for alleged copyright infringement. The suit claims Anthropic fed its LLMs with pirated copies of the authors' work, including from participants Kirk Wallace Johnson, Andrea Bartz and Charles Graeber.^[60]

See also

- Apprenticeship learning
- AI alignment
- Friendly AI
- OpenAI

References

1. Lin, Belle. "Google and Anthropic Are Selling Generative AI to Businesses, Even as They Address Its Shortcomings" (<https://www.wsj.com/articles/google-and-anthropic-are-selling-generative-ai-to-businesses-even-as-they-address-its-shortcomings-ff90d83d>). *WSJ*. Retrieved 2024-04-11.
2. Nellis, Stephen (9 May 2023). "Alphabet-backed Anthropic outlines the moral values behind its AI bot" (<https://www.reuters.com/technology/alphabet-backed-anthropic-outlines-moral-values-behind-its-ai-bot-2023-05-09/>). *Reuters*. Archived (<https://web.archive.org/web/20230605051105/https://www.reuters.com/technology/alphabet-backed-anthropic-outlines-moral-values-behind-its-ai-bot-2023-05-09/>) from the original on 5 June 2023. Retrieved 4 June 2023.
3. Roose, Kevin (11 July 2023). "Inside the White-Hot Center of A.I. Doomerism" (<https://www.nytimes.com/2023/07/11/technology/anthropic-ai-claude-chatbot.html>). *The New York Times*. Archived (<https://web.archive.org/web/20230712235416/https://www.nytimes.com/2023/07/11/technology/anthropic-ai-claude-chatbot.html>) from the original on 12 July 2023. Retrieved 13 July 2023.
4. Palazzolo, Stephanie (8 October 2024). "Anthropic Says Its Chatbot Could Alter Its Hiring Plans" (<https://www.theinformation.com/articles/anthropic-says-its-chatbot-could-alter-its-hiring-plans>). *The Information*.
5. Metz, Cade (2024-03-04). "A.I. Start-Up Anthropic Challenges OpenAI and Google With New Chatbot" (<https://www.nytimes.com/2024/03/04/technology/anthropic-claude-openai-google.html>). *The New York Times*. ISSN 0362-4331 (<https://search.worldcat.org/issn/0362-4331>). Retrieved 2024-03-07.
6. "ChatGPT must be regulated and A.I. 'can be used by bad actors,' warns OpenAI's CTO" (<https://finance.yahoo.com/news/artificial-intelligence-must-regulated-warns-204546351.html>). *Fortune*. 5 February 2023. Archived (<https://web.archive.org/web/20230206075725/https://finance.yahoo.com/news/artificial-intelligence-must-regulated-warns-204546351.html>) from the original on 2023-02-06. Retrieved 2023-02-06 – via finance.yahoo.com.
7. "Research" (<https://www.anthropic.com/research>). *www.anthropic.com*. Retrieved 2024-04-11.
8. "Anthropic | Company Overview & News" (<https://www.forbes.com/companies/anthropic/>). *Forbes*. Retrieved 2024-04-11.

9. Dastin, Jeffrey (September 28, 2023). "Amazon steps up AI race with Anthropic investment" (<https://www.reuters.com/markets/deals/amazon-steps-up-ai-race-with-up-4-billion-deal-invest-anthropic-2023-09-25/>). *Reuters*. Archived (<https://web.archive.org/web/20231105110524/https://www.reuters.com/markets/deals/amazon-steps-up-ai-race-with-up-4-billion-deal-invest-anthropic-2023-09-25/>) from the original on 2023-11-05. Retrieved 2023-10-02.
10. Coldewey, Devin (2024-03-27). "Amazon doubles down on Anthropic, completing its planned \$4B investment" (<https://techcrunch.com/2024/03/27/amazon-doubles-down-on-anthropic-completing-its-planned-4b-investment/>). *TechCrunch*. Retrieved 2024-03-28.
11. Hu, Krystal (October 27, 2023). "Google agrees to invest up to \$2 billion in OpenAI rival Anthropic" (<https://www.reuters.com/technology/google-agrees-invest-up-2-bln-openai-rival-anthropic-wsj-2023-10-27/>). *Reuters*. Archived (<https://web.archive.org/web/20231102122701/https://www.reuters.com/technology/google-agrees-invest-up-2-bln-openai-rival-anthropic-wsj-2023-10-27/>) from the original on 2023-11-02. Retrieved 2023-10-30.
12. Henshall, Will; Perrigo, Billy (2023-09-07). "TIME100 AI 2023: Dario and Daniela Amodei" (<https://time.com/collection/time100-ai/6309047/daniela-and-dario-amodei/>). *Time*. Retrieved 2024-03-05.
13. "As Anthropic seeks billions to take on OpenAI, 'industrial capture' is nigh. Or is it?" (<https://venturebeat.com/ai/as-anthropic-seeks-billions-to-take-on-openai-industrial-capture-is-nigh-or-is-it/>). *VentureBeat*. 2023-04-07. Archived (<https://web.archive.org/web/20230524193007/https://venturebeat.com/ai/as-anthropic-seeks-billions-to-take-on-openai-industrial-capture-is-nigh-or-is-it/>) from the original on 2023-05-24. Retrieved 2023-05-24.
14. "Google invests \$300 million in Anthropic as race to compete with ChatGPT heats up" (<https://venturebeat.com/ai/google-invests-300-million-in-anthropic-as-race-to-compete-with-chatgpt-heats-up/>). *VentureBeat*. 2023-02-03. Archived (<https://web.archive.org/web/20230206075725/https://venturebeat.com/ai/google-invests-300-million-in-anthropic-as-race-to-compete-with-chatgpt-heats-up/>) from the original on 2023-02-06. Retrieved 2023-02-06.
15. Sigalos, MacKenzie (2024-03-25). "FTX estate selling majority stake in AI startup Anthropic for \$884 million, with bulk going to UAE" (<https://www.cnbc.com/2024/03/25/ftx-estate-sells-majority-stake-in-startup-anthropic-for-884-million.html>). *CNBC*. Retrieved 2024-04-11.
16. Perrigo, Billy (2024-05-30). "Inside Anthropic, the AI Company Betting That Safety Can Be a Winning Strategy" (<https://time.com/6980000/anthropic/>). *Time*. Retrieved 2024-06-02.
17. Setty, Riddhi (2023-02-23). "Anthropic A.I. Sues AI Company for Trademark Infringement" (<https://news.bloomberglaw.com/ip-law/anthropic-a-i-sues-ai-company-for-trademark-infringement>). *Bloomberg Law*. Archived (<https://web.archive.org/web/20230222172208/https://news.bloomberglaw.com/ip-law/anthropic-a-i-sues-ai-company-for-trademark-infringement>) from the original on 2023-02-22. Retrieved 2023-02-23.
18. "Amazon and Anthropic announce strategic collaboration to advance generative AI" (<https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai>). *US About Amazon* (Press release). 2023-09-25. Archived (<https://web.archive.org/web/20230925104013/https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai>) from the original on 2023-09-25. Retrieved 2023-09-25.
19. Cosgrove, Emma; Bergman, Ben (November 22, 2024). "Amazon makes massive downpayment on dethroning Nvidia" (<https://www.businessinsider.com/amazon-tries-again-anthropic-ai-chips-trainium-nvidia-2024-11>). *Business Insider*. Retrieved 2024-12-12.
20. Wiggers, Kyle (2024-10-01). "Anthropic hires OpenAI co-founder Durk Kingma" (<https://techcrunch.com/2024/10/01/anthropic-hires-openai-co-founder-durk-kingma/>). *TechCrunch*. Retrieved 2024-10-05.
21. "Our response to the UK Government's internal AI safety policy enquiries" (<https://www.anthropic.com/uk-government-internal-ai-safety-policy-response>). *www.anthropic.com*. Retrieved 2024-04-11.
22. "The Long-Term Benefit Trust" (<https://www.anthropic.com/news/the-long-term-benefit-trust>). *www.anthropic.com*. Retrieved 2024-04-11.

23. Morley, John; Berger, David; Simmerman, Amy (28 October 2023). "Anthropic Long-Term Benefit Trust" (<https://corpgov.law.harvard.edu/2023/10/28/anthropic-long-term-benefit-trust/>). *Harvard Law School Forum on Corporate Governance*. Retrieved 2024-04-03.
24. "Company" (<https://www.anthropic.com/company>). *www.anthropic.com*. Retrieved 2025-04-21.
25. Erin Griffith and Cade Metz: *Inside the Funding Frenzy at Anthropic, One of A.I.'s Hottest Start-Ups*. (<https://www.nytimes.com/2024/02/20/technology/anthropic-funding-ai.html>) In: *NYT*, 20 February 2024. Retrieved 20 February 2024.
26. Pierce, David (May 15, 2024). "Instagram's co-founder is Anthropic's new chief product officer" (<https://www.theverge.com/2024/5/15/24157240/mike-krieger-anthropic-instagram-ai>). *The Verge*.
27. Wiggers, Kyle (May 15, 2024). "Anthropic hires Instagram co-founder as head of product" (<https://techcrunch.com/2024/05/15/anthropic-hires-instagram-co-founder-as-head-of-product/>). *TechCrunch*.
28. Robison, Kylie (May 28, 2024). "OpenAI researcher who resigned over safety concerns joins Anthropic" (<https://www.theverge.com/2024/5/28/24166370/jan-leike-openai-anthropic-ai-safety-research>). *VentureBeat*.
29. Roose, Kevin (2023-10-17). "What if We Could All Control A.I.?" (<https://www.nytimes.com/2023/10/17/technology/ai-chatbot-control.html>). *The New York Times*. ISSN 0362-4331 (<https://search.worldcat.org/issn/0362-4331>). Retrieved 2024-04-11.
30. Drapkin, Aaron (2023-10-27). "What Is Claude AI and Anthropic? ChatGPT's Rival Explained" (<https://tech.co/news/what-is-claude-ai-anthropic>). *Tech.co*. Retrieved 2024-04-11.
31. Deborah, Yao (2023-08-11). "Anthropic's Claude Instant: A Smaller, Faster and Cheaper Language Model" (<https://aibusiness.com/nlp/anthropic-s-claude-instant-a-smaller-faster-and-cheaper-language-model>). *AI Business*. Retrieved 2024-04-09.
32. Sharon, Goldman (2023-03-14). "OpenAI rival Anthropic introduces Claude, an AI assistant to take on ChatGPT" (<https://venturebeat.com/ai/google-funded-anthropic-introduces-claude-chatgpt-rival-through-chat-and-api/>). *Venture Beat*. Retrieved 2024-04-09.
33. Milmo, Dan (12 July 2023). "Claude 2: ChatGPT rival launches chatbot that can summarise a novel" (<https://www.theguardian.com/technology/2023/jul/12/claude-2-anthropic-launches-chatbot-rival-chatgpt>). *The Guardian*. Archived (<https://web.archive.org/web/20231108182005/https://www.theguardian.com/technology/2023/jul/12/claude-2-anthropic-launches-chatbot-rival-chatgpt>) from the original on 8 November 2023. Retrieved 13 July 2023.
34. Matthews, Dylan (17 July 2023). "The \$1 billion gamble to ensure AI doesn't destroy humanity" (<https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2>). *Vox*. Archived (<https://web.archive.org/web/20231003205441/https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2>) from the original on 3 October 2023. Retrieved 23 July 2023.
35. Nuñez, Michael (2024-03-04). "Anthropic unveils Claude 3, surpassing GPT-4 and Gemini Ultra in benchmark tests" (<https://venturebeat.com/ai/anthropic-unveils-claude-3-claims-new-standard-for-intelligence/>). *Venture Beat*. Retrieved 2024-04-09.
36. "Introducing the next generation of Claude" (<https://www.anthropic.com/news/claude-3-family>). Retrieved 4 March 2024.
37. Franzen, Carl (March 4, 2024). "Amazon adds Claude 3, which beats GPT-4, to Bedrock" (<https://venturebeat.com/ai/amazon-adds-claude-3-beats-gpt-4-to-bedrock/>). *Venture Beat*.
38. Field, Hayden (May 1, 2024). "Amazon-backed Anthropic launches iPhone app and business tier to compete with OpenAI's ChatGPT" (<https://www.cnbc.com/2024/05/01/anthropic-iphone-ai-app-business-plan-to-compete-with-openai-announced.html>). *CNBC*. Retrieved May 3, 2024.

39. Pierce, David (2024-06-20). "Anthropic has a fast new AI model — and a clever new way to interact with chatbots" (<https://www.theverge.com/2024/6/20/24181961/anthropic-claude-35-sonnet-model-ai-launch>). *The Verge*. Retrieved 2024-06-22.
40. Nuñez, Michael (2024-10-22). "Anthropic's new AI can use computers like a human, redefining automation for enterprises" (<https://venturebeat.com/ai/anthropic-new-ai-can-use-computers-like-a-human-redefining-automation-for-enterprises/>). *VentureBeat*. Retrieved 2024-11-24.
41. "Palantir IR" (<https://investors.palantir.com/news-details/2024/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations/>). *investors.palantir.com*. Retrieved 2025-02-10.
42. Wiggers, Kyle (2024-12-12). "Anthropic's 3.5 Haiku model comes to Claude users" (<https://techcrunch.com/2024/12/12/anthropics-3-5-haiku-model-comes-to-claude-users/>). *TechCrunch*. Retrieved 2024-12-18.
43. Rooney, Kate (2025-02-24). "Anthropic says it's released its 'most intelligent' AI model yet as competition ramps up" (<https://www.cnbc.com/2025/02/24/anthropic-say-claude-sonnet-3-point7-is-its-most-intelligent-ai-model-yet.html>). *CNBC*. Retrieved 2025-03-17.
44. Heath, Alex (2025-02-24). "Anthropic's new 'hybrid reasoning' AI model is its smartest yet" (<https://www.theverge.com/news/618440/anthropic-claude-3-7-sonnet-ai-model-hybrid-reasoning>). *The Verge*. Retrieved 2025-03-17.
45. Bai, Yuntao; Kadavath, Saurav; Kundu, Sandipan; Askell, Amanda; Kernion, Jackson; Jones, Andy; Chen, Anna; Goldie, Anna; Mirhoseini, Azalia (2022-12-15). "Constitutional AI: Harmlessness from AI Feedback". arXiv:2212.08073 (<https://arxiv.org/abs/2212.08073>) [cs.CL (<https://arxiv.org/archive/cs.CL>)].
46. "Transformer Circuits Thread" (<https://transformer-circuits.pub/>). *transformer-circuits.pub*. Archived (<https://web.archive.org/web/20230204224450/https://www.transformer-circuits.pub/>) from the original on 2023-02-04. Retrieved 2023-02-09.
47. "Towards Monosemanticity: Decomposing Language Models With Dictionary Learning" (<https://transformer-circuits.pub/2023/monosemantic-features/index.html>). Archived (<https://web.archive.org/web/20231009064431/https://transformer-circuits.pub/2023/monosemantic-features/index.html>) from the original on 9 October 2023. Retrieved 10 October 2023.
48. Ropek, Lucas (2024-05-21). "New Anthropic Research Sheds Light on AI's 'Black Box' " (<https://gizmodo.com/new-anthropic-research-sheds-light-on-ais-black-box-1851491333>). *Gizmodo*. Retrieved 2024-05-23.
49. Perrigo, Billy (2024-05-21). "Artificial Intelligence Is a 'Black Box.' Maybe Not For Long" (<https://time.com/6980210/anthropic-interpretability-ai-safety-research/>). *Time*. Retrieved 2024-05-24.
50. Adly Templeton*, Tom Conerly*, Jonathan Marcus, Jack Lindsey, Trenton Bricken, Brian Chen, Adam Pearce, Craig Citro, Emmanuel Ameisen, Andy Jones, Hoagy Cunningham, Nicholas L Turner, Callum McDougall, Monte MacDiarmid, Alex Tamkin, Esin Durmus, Tristan Hume, Francesco Mosconi, C. Daniel Freeman, Theodore R. Sumers, Edward Rees, Joshua Batson, Adam Jermyn, Shan Carter, Chris Olah, Tom Henighan *Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet* (<https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html>), Anthropic, retrieved 24 May 2024
51. Kahn, Jeremy. "Anthropic researchers make progress unpacking AI's 'black box' " (<https://fortune.com/2025/03/27/anthropic-ai-breakthrough-claude-llm-black-box/>). *Fortune*. Retrieved 2025-03-29.
52. Nuñez, Michael (2025-03-27). "Anthropic scientists expose how AI actually 'thinks' — and discover it secretly plans ahead and sometimes lies" (<https://venturebeat.com/ai/anthropic-scientists-expose-how-ai-actually-thinks-and-discover-it-secretly-plans-ahead-and-sometimes-lies/>). *VentureBeat*. Retrieved 2025-03-29.

53. Zeff, Maxwell (2025-01-19). "The Pentagon says AI is speeding up its 'kill chain' " (<https://techcrunch.com/2025/01/19/the-pentagon-says-ai-is-speeding-up-its-kill-chain/>). *TechCrunch*. Retrieved 2025-02-12.
54. Murgia, Madhumita (2024-12-05). "Anthropic's Dario Amodei: Democracies must maintain the lead in AI" (<https://www.ft.com/content/e75e3388-4700-413d-ab67-778410c2d977>). *Financial Times*. Retrieved 2025-02-10.
55. Times, Financial (2023-10-19). "Universal Music sues AI start-up Anthropic for scraping song lyrics" (<https://arstechnica.com/tech-policy/2023/10/universal-music-sues-ai-start-up-anthropic-for-scraping-song-lyrics/>). *Ars Technica*. Retrieved 2024-03-09.
56. David, Emilia (2023-10-19). "Universal Music sues AI company Anthropic for distributing song lyrics" (<https://www.theverge.com/2023/10/19/23924100/universal-music-sue-anthropic-lyrics-copyright-katy-perry>). *The Verge*. Retrieved 2024-03-09.
57. *Court Docket for Concord Music Group v Anthropic* (<https://www.courtlistener.com/docket/67894459/1/concord-music-group-inc-v-anthropic-pbc>), per CourtListener.
58. Masse, Bryson (2023-10-19). "Major music publishers sue Anthropic for copyright infringement over song lyrics" (<https://venturebeat.com/ai/major-music-publishers-sue-anthropic-for-copyright-infringement-over-song-lyrics/>). *VentureBeat*. Retrieved 2024-04-08.
59. Brittain, Blake (2024-01-17). "Anthropic fires back at music publishers' AI copyright lawsuit" (<https://www.reuters.com/legal/litigation/anthropic-fires-back-music-publishers-ai-copyright-lawsuit-2024-01-17/>). *Reuters*. Retrieved 2024-04-07.
60. Field, Hayden (2024-08-20). "Amazon-backed Anthropic hit with class-action lawsuit over copyright infringement" (<https://www.cnbc.com/2024/08/20/amazon-backed-anthropic-hit-with-class-action-lawsuit-over-copyright-infringement.html>). *CNBC*. Retrieved 2024-10-29.

External links

- [Official website \(<https://www.anthropic.com/>\)](https://www.anthropic.com/) 
-

Retrieved from "<https://en.wikipedia.org/w/index.php?title=Anthropic&oldid=1288694163>"